

**TMDB ICON**

**Raccomandazione di film basata sulla comunità di TMDB**

**Membri del gruppo:**

Spicoli Piersilvio 736519

Papagno Francesco 735596

Brugnola Danilo 745960

Progetto disponibile al seguente link:

Sommario

[INTRODUZIONE 2](#_Toc150334072)

[STEP 1 3](#_Toc150334073)

# INTRODUZIONE

L'industria dei servizi online e del commercio elettronico ha sperimentato una rapida crescita, portando a una vasta quantità di dati e prodotti disponibili per gli utenti. In questo contesto, i sistemi di raccomandazione hanno acquisito crescente importanza, fornendo suggerimenti personalizzati agli utenti per migliorare l'esperienza di navigazione e l'engagement. La raccomandazione è una delle task chiave nell'ambito dell'apprendimento automatico, mirata a prevedere le preferenze degli utenti su elementi come prodotti, film, musica o contenuti digitali.

L'attuale **stato dell'arte** nel campo dei modelli di raccomandazione si basa ampiamente su tecniche avanzate di machine learning. Tra queste, i modelli di **Collaborative Filtering (CF)** sfruttano i pattern di comportamento degli utenti o delle entità simili per fare previsioni (U2U e I2I), mentre le tecniche di **Content-Based Filtering (CBF)** analizzano le caratteristiche degli elementi raccomandati e i profili degli utenti per generare raccomandazioni personalizzate.

La domanda che noi membri del team ci siamo posti è la seguente:

*è possibile realizzare e implementare un task di raccomandazione usando strategie ben diverse da quelle offerte dallo stato d’arte?*

Il presente progetto propone un approccio innovativo alla task di raccomandazione **attraverso un processo di inferenza guidato da algoritmi di machine learning regressori**. Piuttosto che affidarsi agli standard dell'attuale stato dell'arte dei modelli di raccomandazione, il nostro approccio si basa sull'inferenza dei gusti e delle preferenze degli utenti (che appartengono nel nostro caso alla community di TMDB) utilizzando algoritmi di regressione. Questo approccio permette una comprensione più profonda e granulare dei comportamenti degli utenti, consentendo di offrire raccomandazioni altamente personalizzate e adattabili. E, oltre questo, per comprendere appieno le potenzialità offerte dalle tecniche implementate da noi stessi.

In questo documento verranno illustrate, pertanto, le fasi per la realizzazione del progetto TMDB ICON relativo al corso di studi Ingegneria della Conoscenza a.a. 2022/2023. Per comprendere appieno il problema di raccomandazione, abbiamo dovuto riformulare il problema come se fosse un

**Task 1**

Raccomandare i primi 10 film che hanno maggior ranking di gradimento per la comunità di utenti di TMDB utilizzando una strategia basata su inferenze

**Task 2**

Raccomandare i primi 10 film che hanno maggior ranking di gradimento per la comunità di utenti di TMDB utilizzando una strategia basata COP (Costraint Optimization Problem)

Prima dell’esecuzione del codice è necessaria l’installazione delle librerie contenute nel file di testo *requirements.txt* utilizzando il seguente comando:



Inoltre, bisognerà installare la versione di *SWI Prolog 9.0.4.*

Durante l'esecuzione potrebbero comparire dei warning dettati dalla esecuzione del codice, non sono degli errori bloccanti in quanto il codice è stato testato ed i grafici posti all'interno della documentazione sono derivanti dall'esecuzione del codice.

Per completare le due task, la realizzazione è stata divisa in tre fasi principali, ossia:

* Rappresentazione
* Ragionamento
* Apprendimento

# STEP 1

Nel primo step andiamo a rappresentare il mondo di riferimento seguendo un modello comprensibile alla macchina in modo da permettergli l'elaborazione dei dati che andremo ad introdurre in essa.

Nel nostro caso, gli elementi che descrivono il nostro dominio vengono descritti tramite “features”, ossia variabili che denotano le caratteristiche del nostro mondo.

Le features con la quale abbiamo lavorato per completare le nostre task sono state fornite da due dataset: *tmdb\_5000\_credits.csv* e *tmdb\_5000\_movies.csv.* Questi due dataset sono stati presi dal sito web *Kaggle*.

Dopo aver scaricato i dataset viene eseguita la fase di preprocessing descritta di seguito:

**ANALISI DEL DATASET TROVATO E MODELLAZIONE**

Il dataset: *tmdb\_5000\_credits.csv* contiene le seguenti features:

* *movie\_id:* contiene il codice ID che individua univocamente il film
* *title:* contiene il titolo del film
* *cast:* contiene diverse informazioni riguardanti il cast del film
* *crew:* contiene diverse informazioni riguardanti la crew di produzione del film

Mentre, il dataset *tmdb\_5000\_movies.csv* contiene le seguenti features:

* *budget:* contiene il prezzo del budget utilizzato per realizzare il film
* *genres:* contiene i generi del film
* *hompage:* contiene il sito web del film
* *id:* contiene l’ID del film
* *keywords:* contiene alcune parole chiave inerenti al film
* *original\_language:* contiene la lingua originale del film
* *original\_title:* contiene il titolo originale del film
* *overview:* contiene la trama del film
* *popularity:* contiene un valore che indica la popolarità del film
* *production\_companies:* contiene le compagnie di produzione del film
* *production\_countries:* contiene i continenti nella quale il film è stato prodotto
* *release\_date:* contiene la data di rilascio de film
* *revenue:*
* *runtime:* contiene la durata del film indicata in minuti
* *spoken\_languages:* contiene le lingue che vengono parlate nel film
* *status:* indica se il film è stato rilasciato o no
* *tagline:* contiene un ulteriore tag del film
* *title:* contiene il titolo del film in lingua inglese
* *vote\_average:* indica la media dei voti attribuiti al film
* *vote\_count:* indica il numero di voti che il film ha ricevuto

Partendo da questo dataset iniziale e considerando le nostre due task, abbiamo generato #numero dataset distinti droppando alcuni features che non erano rilevanti (elenco features droppate).

I dataset ottenuti sono: